CLAIMS

What is claimed is:

1. A method of producing a model that predicts the lability of reactive sites on a chemical compound, the method comprising:

(a) obtaining structural representations for a training set of chemical compounds;

(b) for each of said chemical compounds, identifying one or more reactive sites pertinent to the model;

(c) for each of said reactive sites,

(i) obtaining a lability value from a trustworthy source or technique; and

(ii) characterizing the reactive site in terms of values for a plurality of chemical structural descriptors, including at least two of an atom type at the reactive site, atoms types at neighboring positions to the reactive site, a partial charge on an atom or group at the reactive site, and a geometric characterization of the reactive site; and

(d) for all of said reactive sites, using the lability values and chemical structural descriptor values to obtain an expression for lability that sums contributions from each of the chemical structural descriptors.

2. The method of claim 1, wherein the structural representations are three dimensional depictions including at least bond lengths and bond angles.

3. The method of claim 1, wherein the model predicts lability for an oxidation reaction selected from the group consisting of aromatic oxidation, aliphatic hydrogen atom abstraction, carbon-carbon double bond oxidation, nitrogen atom oxidation, and sulfur atom oxidation.

4. The method of claim 3, wherein identifying reactive sites pertinent to the model comprises identifying sites where the oxidation reaction can occur on the chemical compounds of the training set.

5. The method of claim 1, wherein obtaining a lability value from a trustworthy technique comprises calculating at least one of an activation energy, an ionization potential of an intermediate radical formed at a reactive site under consideration, and a delta enthalpy of formation for the intermediate radical formed at the reactive site under consideration versus the base chemical compound under consideration.

6. The method of claim 5, wherein the trustworthy technique employs a quantum mechanical representation of the chemical compounds of the training set.

7. The method of claim 1, wherein the model is an aliphatic hydrogen abstraction model and wherein the descriptors comprise fragment-based descriptors or geometry-based descriptors.

8. The method of claim 1, wherein the model is an aromatic oxidation model, and wherein descriptors comprise fragment-based descriptors or geometry-based descriptors.

9. The method of claim 1, wherein using the lability values and the chemical structural descriptor values to obtain an expression for lability comprises employing a regression technique.

10. The method of claim 9, wherein the regression technique is selected from the group consisting of partial least squares, principal component analysis, and linear regression techniques.

11. A method implemented on a computing device for predicting labilities of reactive sites on a chemical compound, the method comprising:

(a) identifying a reactive site on the chemical compound;

(b) identifying values for a plurality of chemical structural descriptors for the reactive site, said chemical structural descriptors specifying at least one of an atom type at the reactive site, atom types at neighboring positions to the reactive site, a partial charge on the atom or group at the reactive site, and a geometric characterization of the reactive site;

(c) calculating a lability value for the reactive site by summing terms of an expression, wherein the terms include or are derived from individual ones of the chemical structural descriptors;

(d) repeating (a) – (c) for more additional reactive sites of the chemical compound; and

(e) outputting lability values calculated at (c) for the reactive sites on the chemical compound.

12. The method of claim 11, wherein the reactive site is susceptible to one of the following oxidation reactions: aliphatic hydrogen atom abstraction, aromatic oxidation, carbon-carbon double bond oxidation, nitrogen oxidation, and sulfur oxidation.

13. The method of claim 12, wherein the chemical structural descriptors are specific for the oxidation reaction to which the reactive site is susceptible.

14. The method of claim 11, wherein the chemical structural descriptors are selected from the group consisting of an atom type of the reactive site, atom types at neighboring positions of the reactive site, a partial charge on an atom or group at the reactive site, a geometric characterization of the reactive site, and combinations thereof.

15. The method of claim 11, wherein the expression is a linear expression for the lability value and having a defined coefficient for each chemical structural descriptor in the expression.

16. The method of claim 12, wherein the expression is specific for the oxidation reaction to which the reactive site is susceptible.

17. The method of claim 11, wherein the lability value represents an activation energy of an oxidation reaction at the reactive site, an ionization potential of an intermediate radical generated by the oxidation reaction at the reactive site, or a delta enthalpy of formation of the intermediate radical formed by the oxidation reaction at the reactive site.

18. The method of claim 11, further comprising simultaneously displaying the lability values calculated at (c) for all reactive sites.

19. The method of claim 11, wherein outputting the lability values calculated at (c) comprises sending the lability values to a remote network site from which a request to predict lability values has originated.

20. The method of claim 11, further comprising correcting the lability values calculated at (c) by modifying said lability values to account for one or more particular reaction characteristics of one or more cytochrome P450 enzymes.

21. A computer program product comprising a machine readable medium on which is provided programmed instructions for producing a model that predicts the lability of reactive sites on a chemical compound, the programmed instructions comprising:

(a) program code for obtaining structural representations for a training set of chemical compounds;

(b) for each of said chemical compounds, program code for identifying one or more reactive sites pertinent to the model;

(c) for each of said reactive sites,

(i) program code for obtaining a lability value from a trustworthy source or technique; and

(ii) program code for characterizing the reactive site in terms of values for a plurality of chemical structural descriptors, including at least two of an atom type at the reactive site, atoms types at neighboring positions to the reactive site, a partial charge on an atom or group at the reactive site, and a geometric characterization of the reactive site; and

(d) for all of said reactive sites, program code for using the lability values and chemical structural descriptor values to obtain an expression for lability that sums contributions from each of the chemical structural descriptors.

22. The computer program product of claim 21, wherein the structural representations are three dimensional depictions including at least bond lengths and bond angles.

23. The computer program product of claim 21, wherein the model predicts lability for an oxidation reaction selected from the group consisting of aromatic, oxidation, aliphatic hydrogen atom abstraction, carbon-carbon double bond oxidation, nitrogen atom oxidation, and sulfur atom oxidation.

24. The computer program product of claim 23, wherein the program code for identifying reactive sites pertinent to the model comprises program code for identifying sites where the oxidation reaction can occur on the chemical compounds of the training set.

25. The computer program product of claim 21, wherein the program code for obtaining a lability value from a trustworthy technique comprises program code for calculating at least one of an activation energy, an ionization potential of an intermediate radical formed at a reactive site under consideration, and a delta enthalpy of formation for the intermediate radical formed at the reactive site under consideration versus the base chemical compound under consideration.

26. The computer program product of claim 25, wherein the trustworthy technique employs a quantum mechanical representation of the chemical compounds of the training set.

27. The computer program product of claim 21, wherein the program code for using the lability values and the chemical structural descriptor values to obtain an expression for lability comprises program code employing a regression technique.

28. The computer program product of claim 7, wherein the regression technique is selected from the group consisting of partial least squares, principal component analysis, and linear regression techniques.

29. A computer program product comprising a machine readable medium on which is provided programmed instructions for predicting labilities of reactive sites on a chemical compound, the programmed instructions comprising:

(a) program code for identifying a reactive site on the chemical compound;

(b) program code for identifying values for a plurality of chemical structural descriptors for the reactive site, said chemical structural descriptors specifying at least one of an atom type at the reactive site, atom types at neighboring positions to the reactive site, a partial charge on the atom or group at the reactive site, and a geometric characterization of the reactive site;

(c) program code for calculating a lability value for the reactive site by summing terms of an expression, wherein the terms include or are derived from individual ones of the chemical structural descriptors;

(d) program code for repeating (a) – (c) for more additional reactive sites of the chemical compound; and

(e) program code for outputting lability values calculated at (c) for the reactive sites on the chemical compound.

30. The computer program product of claim 29, wherein the reactive site is susceptible to one of the following oxidation reactions: aliphatic hydrogen atom abstraction, aromatic oxidation, carbon-carbon double bond oxidation, nitrogen oxidation, and sulfur oxidation.

31. The computer program product of claim 30, wherein the chemical structural descriptors are specific for the oxidation reaction to which the reactive site is susceptible.

32. The computer program product of claim 29, wherein the chemical structural descriptors are selected from the group consisting of an atom type of the reactive site, atom types at neighboring positions of the reactive site, a partial charge on an atom or

group at the reactive site, a geometric characterization of the reactive site, and combinations thereof.

33. The computer program product of claim 29, wherein the expression is a linear expression for the lability value and having a defined coefficient for each chemical structural descriptor in the expression.

34. The computer program product of claim 30, wherein the expression is specific for the oxidation reaction to which the reactive site is susceptible.

35. The computer program product of claim 29, wherein the lability value represents an activation energy of an oxidation reaction at the reactive site, an ionization potential of an intermediate radical generated by the oxidation reaction at the reactive site, or a delta enthalpy of formation of the intermediate radical formed by the oxidation reaction at the reactive site.

36. The computer program product of claim 29, further comprising program code for simultaneously displaying the lability values calculated at (c) for all reactive sites.

37. The computer program product of claim 29, wherein the program code for outputting the lability values calculated at (c) comprises program code for sending the lability values to a remote network site from which a request to predict lability values has originated.

38. The computer program product of claim 29, further comprising program code for correcting the lability values calculated at (c) by modifying said lability values to account for one or more particular reaction characteristics of one or more cytochrome P450 enzymes.

39. A method for calculating a linear regression equation with a set of organic chemical descriptors, each descriptor having a coefficient, the method comprising:

(a) identifying the reactive sites on each substrate molecule in a training set of substrate molecules;

(b) obtaining activation energy and reactivity values for each of the reactive sites from an external method;

(b) determining the organic chemical descriptors that describe each of the reactive sites; and

(c) calculating the coefficients and the linear regression equation.

40. The method of claim 39 wherein the organic chemical descriptors comprise site atom descriptors and neighbor atom descriptors.

41. The method of claim 39 wherein the organic chemical descriptors comprise partial charges, total charge and bond length descriptors.

42. The method of claim 39 wherein the linear regression equation is used to model and predict the reactivity of other substrate molecules.

43. The method of claim 42 wherein the linear regression equation is used to model and predict the reactivity of other substrate molecules in cytochrome p450 metabolism.

44. The method of claim 42 wherein the substrate molecules are drug candidates.

45. A method for predicting the reactivity of a substrate molecule, the method comprising:

(a) identifying reactive sites on the substrate molecule;

(b) characterizing the reactive sites based on organic chemical descriptors;

(c) using the organic chemical descriptors in a linear regression equation designed to model and predict the reactivity of substrate molecules, in order to predict the reactivity of the substrate molecule;

wherein the organic chemical descriptors are the same descriptors used to derive the linear regression equation from a training set of substrate molecules.

46. The method of claim 45 wherein the organic chemical descriptors comprise site atom descriptors and neighbor atom descriptors.

47. The method of claim 45 wherein the organic chemical descriptors comprise partial charges, total charge and bond length descriptors.

48. The method of claim 45 wherein the reactivity of the substrate molecule is calculated specifically with respect to the cytochrome p450 metabolic cycle.

49. The method of claim 45 wherein the substrate molecules are drug candidates.

50. The method of claim 45 wherein the predicted reactivity of a substrate molecule is adjusted with a steric correction factor.

51. A computer program product comprising a machine readable medium on which is stored program code for calculating a linear regression equation with a set of organic chemical descriptors, each descriptor having a coefficient, the program code specifying instructions for:

(a) identifying the reactive sites on each substrate molecule in a training set of substrate molecules;

(b) obtaining activation energy and reactivity values for each of the reactive sites from an external method;

(b) determining the organic chemical descriptors that describe each of the reactive sites; and

(c) calculating the coefficients and the linear regression equation.

52. A computer program product comprising a machine readable medium on which is stored program code for creating a training set of substrate molecules, each substrate molecule comprising at least one reactive site, and each reactive site comprising a plurality of organic chemical descriptors.

53. A computer program product comprising a machine readable medium on which is stored program code for predicting the reactivity of a substrate molecule, each descriptor term having a coefficient, the program code specifying instructions for:

(a) identifying reactive sites on the substrate molecule;

(b) characterizing the reactive sites based on organic chemical descriptors;

(c) using the organic chemical descriptors in a linear regression equation designed to model and predict the reactivity of substrate molecules, in order to predict the reactivity of the substrate molecule;

wherein the organic chemical descriptors are the same descriptors used to derive the linear regression equation from a training set of substrate molecules.